

Safe Markov Chains for Density Control of ON/OFF Agents with Observed Transitions

Nazlı Demirer, Mahmoud El Chamie, and Behçet Açıkmüşe

Abstract—This paper presents a convex optimization approach to control the density distribution of autonomous mobile agents with two control modes: ON and OFF. The main new characteristic distinguishing this model from standard Markov decision models is the existence of the ON control mode and its observed actions. When an agent is in the ON mode, it can measure the instantaneous outcome of one of the actions corresponding to the ON mode and decides whether it should take this action or not based on this new observation. If it does *not* take this action, it switches to the OFF mode where it transitions to the next state based on a predetermined set of transitional probabilities, without making any additional observations. In this decision-making model, each agent acts autonomously according to an ON/OFF decision policy, and the discrete probability distribution for the agent's state evolves according to a discrete-time Markov chain that is a linear function of the stochastic environment (i.e., transition probabilities) and the ON/OFF decision policy. The relevant policy synthesis is formulated as a convex optimization problem where hard safety and convergence constraints are imposed on the resulting Markov matrix. We first consider the case where the ON mode has a single action and the OFF mode has deterministic transitions (rather than stochastic) to demonstrate the model and the ideas behind our approach, which is then generalized to the case where ON mode has multiple actions and OFF mode has stochastic transitions.

I. INTRODUCTION

This paper presents a convex optimization based approach for the synthesis of randomized decision policies to control the density of mobile agents that switch between two modes: ON and OFF. Each mode consists of a (possibly overlapping) finite set of actions, that is, there exist a set of actions for the ON mode and another set for the OFF mode, which may have a non-empty intersection. At each time step, an agent is allowed to measure (observe) the instantaneous outcome, i.e. *transition*, for a single action it chooses among the set of actions for the ON mode (the actions that can be taken while in ON mode). Then, it decides whether to accept or reject this transition. Both decisions, i.e., selection of an action to observe and acceptance/rejection of corresponding transition are made based on a randomized decision policy. If the agent does not accept the proposed transition, it switches to the OFF mode where it transitions to the next state based on a predetermined set of transitional probabilities, without making any additional observations. The probability distribution of each agent's state evolves according to the resulting, underlying, finite-state and discrete-time Markov chain (MC). In the example of large number of multiple, *swarm of*, autonomous agents, the overall density distribution of the swarm can be controlled by designing the underlying MC via the ON/OFF decision policy.

Authors are with the Department of Aeronautics and Astronautics, University of Washington, Seattle, WA 98195. Emails: {ndemirer, melchami, behcet}@uw.edu

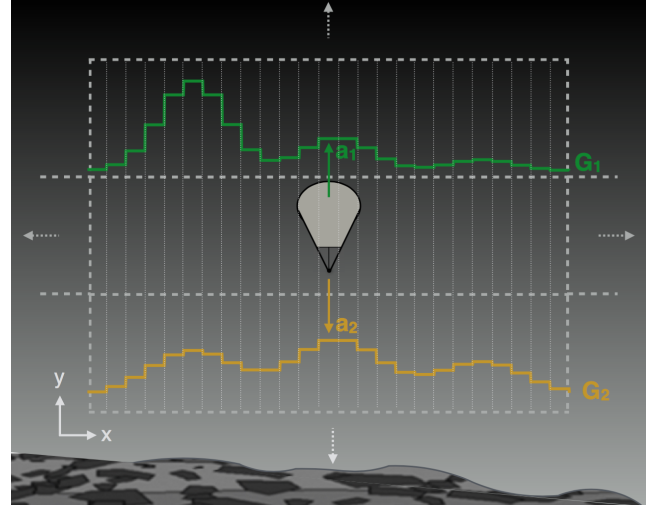


Fig. 1. A simple 2D air balloon illustration with actions, a_i . G_i is the discrete probability density distribution for the x -position of the balloon resulting from taking action a_i from its current position. In this example, the balloon observes outcome of an action by changing its altitude based on the action, and the OFF mode is an MDP [11] running over all actions, hence G_{off} is a function of G_i 's, i.e., $G_{\text{off}} = f(G_1, G_2, \dots)$.

The control of systems with autonomous mobile agents has been a point of interest recently in many applications involving multi-agent sensor networks [1] such as micro vehicles for surveillance, search and rescue, mobile space-based [2], [3], aerial [4], [5], naval [6], or land-based sensor networks [7]. The proposed Markov decision model and formulation is applicable for systems with both single and multiple agents, i.e., the density distribution can be interpreted as the temporal probability distribution of the state of a single agent, or the state probability distribution over multiple agents. Among many possible applications of the theoretical framework presented, we illustrate the results with a swarm control example. Our previous research has developed methods for swarm control policy synthesis without the notion of agent modes or actions [8], [3], [9], then the idea is extended to control of a swarm of partially controlled ON/OFF agents in [10], where we only considered the case when there is only a single action for the ON mode and a deterministic action for the OFF mode. In this paper, we generalize this model for the case when there is an ON mode with multiple actions and an OFF mode with stochastic transitions, through a new Markov decision model with additional measurements for state transitions.

It is noteworthy that the OFF mode captures multiple interesting scenarios: (i) There is only a single action for the OFF mode for which the outcome is not observable; (ii) There are multiple actions for the OFF mode without outcome observations and a standard decision policy running over these

actions, which results in an effective transition matrix for the underlying Markov chain; iii) Transitions can be observed for all available actions and there is a standard decision policy that can run over the actions without requiring transition observations, which is utilized as the default policy when the observed action is rejected. Note that in the third interpretation, the set of actions for the ON mode and the OFF mode are identical, and a rejected observed action at some instant may be still chosen when the default OFF mode decision policy is executed after rejection.

For the formulation of the ON/OFF policy synthesis problem, we define two decision variables for each action: i) probability of choosing an action, i.e., a mapping from states to a probability distribution over the finite number actions; ii) probability of accepting the corresponding transition with the chosen action. The policy synthesis problem turns out to be bilinear (hence non-convex) in these decision variables. However, applying a change of variable, we show that we can still obtain optimal solutions by solving an equivalent convex linear matrix inequality (LMI) optimization problem. When implementing this method, the decision policies are computed offline by solving LMI problem and given to each agent assuming that each agent can observe its current state, measure one-step outcome of a single action for the ON mode, and accept or reject a transition corresponding to the selected action at each time step. If this action is rejected, the agent switches to the OFF mode. In this sense ON and OFF modes can be seen as higher level actions, under which there are the lower level motion actions.

An interesting example to illustrate the concept of ON/OFF agents is controlling air balloons in uncertain wind fields [5] for scientific measurements in Earth and other planets [12], [13]. Probabilistic wind velocity field information, i.e., the probability density functions for the wind speed and direction, changes as a function of the balloon's altitude. The hot air balloons can change their altitudes to choose the velocity field that they ride with, i.e., horizontal motion induced by the wind. Previous research proposed controlled Markov process models for this example [5]. In our approach, the action in this Markov process model is the choice of altitude (see Figure 1). Then we can synthesize a default Markov Decision Process (MDP) policy to distribute the balloons based on this prior wind field knowledge (transition probabilities as a function of the altitude), which defines the OFF mode. We can also design an ON policy, for which we measure the instantaneous velocity observed at different altitudes. We can accept or reject the selected altitude (action) for the ON mode based on the instantaneous velocity observed at this altitude. If it is rejected based on the observed outcome, we can go to the altitude suggested by the OFF mode policy. Note that, the sets of actions for the ON and OFF modes are the same in this example.

II. RELATED RESEARCH

The proposed Markov model is applicable to both decision-making for single and multi-agent systems in stochastic environments. Our particular interest is motivated by the problem

of guiding a multi-agent system, which has been a recent subject of research in Markov Decision Processes (MDPs) [14], [15], [16]. The problem presented here can also be considered as an MDP [11] with finite number of states and actions, and a new set of observations. However, rather than having a reward function for each action and transition [17], the mission objectives here are embedded within the underlying Markov chain.

Most of the work in the area of MDPs and multi-agent systems focuses on the decentralized control of Markov decision processes (Dec-MDP) and decentralized partially-observed Markov decision processes (Dec-POMDP), e.g., [18], where each agent has partial or incomplete observations of its state. These problems, generally, are very difficult to solve and become intractable for large scale problems [19]. These problems with incomplete observations are quite interesting, but not directly comparable to the problem considered in this paper, where we utilize additional measurements for state-transitions.

Using a Markov chain for guiding large numbers of agents to a desired distribution is a relatively new idea. [20] considers a similar problem in the task allocation framework where probability of switching between tasks are designed to achieve maximum redistribution rate, without any additional constraints. A different Markov chain based method is proposed in [21] by using a probabilistic disablement approach found in [22]. [23] uses biased random walk, which leads vehicle positions to evolve toward a probability density. Another method for multi-agent coordination is to use nearest neighbor information to establish consensus [24], [25], [26], [27], [28]. [29] proposes a method based on locational optimization and centroidal Voronoi diagrams, while [30], [31] provide adaptive methods of multi-agent coordination. Other stochastic approaches to multi-agent control include gradient-based decentralized controllers [32] utilizing relative positions to neighbors; game theoretical formulations [33], where each vehicle is considered as a self-interested player; maze searching techniques [34] and cyclic pursuit strategies [35], where each agent pursues its leading neighbor resulting in convergence to a prescribed geometric pattern. In [36], a quorum based method is proposed where the multi-agent system is modeled as a hybrid system. Reference [37] presents a comprehensive survey of recently developed theoretical tools for modeling, analysis, and design of motion coordination algorithms. An interesting application introduced in the previous section, which can also be explored in the proposed ON/OFF framework, is the control of hot air balloons in stochastic wind fields for atmospheric science observations [13]. Earlier work converted this motion planning problem into a more standard Markov decision model [12], [5]. The main distinction of our paper from the listed references above is the existence of the ON control mode and its observed actions. This allows us to devise new methods to control the density distribution of autonomous agents via a new Markov decision model with measurements on the state transitions. Measurements for the ON mode can be obtained by the deployment of additional sensors to extend the agents' sensing capabilities. In summary, the key contributions of this paper are: i) Formulation of a

new Markov chain synthesis problem through a new Markov decision model, with additional measurements for the state transitions, where a policy is designed to ensure that the desired safety and convergence properties for the underlying Markov chain; ii) Convexification of the synthesis problem; iii) Application of the model to density control of swarm of autonomous mobile agents.

The rest of the paper is organized as follows: Section III summarizes the formulation of the density control problem for ON/OFF agents for the case where ON mode has a single action and OFF mode has deterministic transitions; Section IV generalizes the problem to having multiple actions for the ON mode and stochastic transitions for the OFF mode, provides the algorithm for implementation and makes connections to MDPs; Section V describes the ergodicity, transition and safety constraints and formulates the convex LMI problem with this constraints and Section VI has an illustrative example which uses the Markov matrix with the safety constraints. Section VII concludes the paper.

Notation

The following is a partial list of notation used: $\mathbf{0}$ is the zero vector/matrix of appropriate dimensions; I is the identity matrix; $\mathbf{1}$ is the vector of ones with appropriate dimensions; e_i is a vector of appropriate dimension with its i^{th} entry +1 and its other entries zeros; $x[i] = e_i^T x$ for any $x \in \mathbb{R}^n$, and $A[i, j] = e_i^T A e_j$ for any $A \in \mathbb{R}^{n \times m}$; $Q = Q^T \succ (\succeq) 0$ implies that Q is a symmetric positive (semi-)definite matrix; $R \succ (\succeq) H$ implies that $R[i, j] \succ (\succeq) H[i, j]$ for all i, j ; $R \succ (\succeq) \mathbf{0}$ implies that R a positive (non-negative) matrix; $x \in \mathbb{P}^n$ is said to be a *probability vector* if $x \geq \mathbf{0}$ and $\mathbf{1}^T x = 1$; matrix $M \in \mathbb{P}^{m \times m}$ is a Markov matrix if $M \geq \mathbf{0}$ and $\mathbf{1}^T M = \mathbf{1}^T$; \mathcal{P} denotes probability of a random variable; \mathbb{R}^n is the n dimensional real vector space; \mathbb{N} is set of nonnegative integers, i.e., $\mathbb{N} = \{0, 1, 2, \dots\}$; $\mathbb{N}_n^+ = \{1, 2, \dots, n\}$; \emptyset denotes the empty set; (v_1, v_2, \dots, v_n) represents a vector obtained by augmenting vectors v_1, \dots, v_n such that $(v_1, v_2, \dots, v_n) \equiv [v_1^T \ v_2^T \ \dots \ v_n^T]^T$ where v_i can have arbitrary dimensions; $\text{diag}(A) = (A[1, 1], \dots, A[n, n])$ for matrix A ; $\text{diag}(v)$ is a square diagonal matrix with the elements of vector v on the main diagonals; \otimes denotes the Kronecker product; \odot represents the Hadamard (Schur) product; $\mathbf{i}(A)$ is the indicator matrix for any matrix A , whose entries are given by $\mathbf{i}(A)[i, j] = 1$ if $A[i, j] \neq 0$ and $\mathbf{i}(A)[i, j] = 0$ otherwise. $\eta \sim U(0, 1)$ denotes a random variable sampled from the uniform distribution in the interval $[0, 1]$.

III. MARKOV CHAIN MODEL FOR DENSITY CONTROL OF ON/OFF AGENTS: SINGLE ON ACTION CASE

In this section, we introduce the probabilistic density control problem for ON/OFF agents. First, we consider the simplified model of mode-switching ON/OFF agents to demonstrate the concept and formulation that first appeared in [10]. Then, we generalize this model in the next section. In both general and simplified models, agents are assumed to know their current states. For the simplified model, we consider the case where the ON mode contains only one action for which the action

outcome is observable, i.e., agents can measure their next state at a given instant if they take the action (i.e., they decide to be in the ON mode). However agents do not have this transitional observation for the single action in the OFF mode. For ease of demonstration of the idea, we assume that the stochastic transition matrix for the action in OFF mode is an identity matrix. Note that having an identity matrix means that the outcomes of the action for the OFF mode are deterministic (they do not require measurement of the transitions), which is a trivial case and it will be generalized much further in the next section. This assumption makes the initial formulation more transparent. Furthermore, this simpler model also captures an interesting interpretation of the density control problem as presented in [10], that is, ON/OFF agents are moving with the dynamics induced by the stochastic environment when they are ON and they are staying stationary when they are OFF.

The density control problem for this simplified model can be formulated as a Markov chain synthesis problem. For that, we define a finite set of states $\mathcal{S} = \{s_1, \dots, s_n\}$, that is, there is a discrete state space with cardinality n and s_j is referred to as “ j^{th} state” in remainder of the paper. We consider a discrete-time system where $s(t) \in \mathcal{S}$ is the state of the agent at time epoch t , i.e., $s(t) = s_i$ is the event that the state is the i^{th} state at time t . Then, probability density distribution $x(t) \in \mathbb{P}^n$ is defined as:

$$x[i](t) = \mathcal{P}\{s(t) = s_i\}, \quad i \in \mathbb{N}_n^+, \quad t \in \mathbb{N}, \quad (1)$$

where t is the discrete time index. Hence $x[i](t)$ is the probability of a mobile agent to be in the i^{th} state at time t . In the rest of the section, we present the formulation for the time evolution of the density $x(t)$ as the following Markov chain, which is defined over the state-space \mathcal{S} :

$$x(t+1) = Mx(t), \quad (2)$$

where M is the transition matrix, i.e., $M[i, j] = \mathcal{P}\{s(t+1) = s_i | s(t) = s_j\}$. The transition matrix M is a function of the stochastic environment and the ON/OFF policy, as will be explained next. We define the following events to properly define the stochastic environment and the decision policy, for $t \in \mathbb{N}$:

- $y(t+1) = s_l$: Observing a transition to state s_l
- $\sigma(t) = \sigma_{on}$: Accepting to execute the action for the ON mode.

Note that we have two *modes* of operation $\sigma(t) \in \{\sigma_{on}, \sigma_{off}\}$. Even though $y(t+1)$ has a time index $t+1$, this observation occurs at time t . In particular, observation of a transition is different from the actual transition taking place. For example, $y(t+1) = s_l$ is the event that the stochastic environment would have caused a transition to l^{th} state at time $t+1$ if the action for the ON mode were to be accepted at time t (i.e., observing one-step ahead in the future), whereas $s(t+1) = s_l$ is the event that the transition to l^{th} state has actually occurred.

The stochastic environment is defined with the transition matrix $G \in \mathbb{P}^{n \times n}$, where $G[i, j]$ is the probability of observing a transition from j^{th} state to i^{th} state when the action for the

ON mode is taken, i.e.,

$$G[i, j] := \mathcal{P}\{y(t+1) = s_i | s(t) = s_j\}. \quad (3)$$

In this paper, the environment transition matrices and the decision policy are assumed to be time-invariant (i.e., the processes are stationary), hence the corresponding Markov chain transition matrix given in (2) is also time-invariant. Our objective is to synthesize a decision policy for an agent to accept or reject the corresponding transition observed for the action in the ON mode at each time epoch, i.e. to decide whether it should be ON or OFF, such that the resulting Markov chain will satisfy the desired transition and safety constraints while guiding the density distribution to a desired final distribution. In this section, the state of the agent is assumed not to change if it is OFF, i.e., $s(t+1) = s(t)$ when $\sigma(t) = \sigma_{\text{off}}$. Then the probabilistic ON/OFF decision policy is defined by a matrix $K \in \mathbb{R}^{n \times n}$ (to be designed) that satisfies:

$$0 \leq K \leq \mathbf{1}\mathbf{1}^T, \quad \text{diag}(K) = \mathbf{1}, \quad (4)$$

where $K[i, j]$ is the probability of being ON, given that the transition from j^{th} state to i^{th} state is observed, i.e., the acceptance probability of the environmentally induced motion determined by G :

$$K[i, j] = \mathcal{P}\{\sigma(t) = \sigma_{\text{on}} | s(t) = s_j, y(t+1) = s_i\}. \quad (5)$$

Note that for the transition $i \rightarrow i$, $\forall i$, accepting or rejecting the transition corresponds to the same outcome in this case (in this section). Hence, the diagonal elements of acceptance matrix K are set to 1. We can now give the resulting ON/OFF decision-making policy for an agent in Algorithm 1.

Algorithm 1: ON/OFF Decision-Making Policy – Single action case

Inputs: K (matrix designed offline), $\mathcal{S}, t_{\text{max}}$

```

1 for  $t \leftarrow 1$  to  $t_{\text{max}}$  do
2   Observe the current state,  $s(t) \in \mathcal{S}$  (assume  $s(t) = s_j$ );
3   Generate a random number  $\eta(t) \sim U(0, 1)$ ;
4   Observe the transition outcome if the agent is in the ON mode, i.e.,  $y(t+1)$  (assume  $y(t+1) = s_i$ );
5   if  $\eta(t) \in [0, K[i, j]]$  then
6     The agent switches to the mode ON
      $\sigma(t) = \sigma_{\text{on}}$ , and  $s(t+1) = y(t+1)$ ;
7   else
8     The agent switches to the mode OFF
      $\sigma(t) = \sigma_{\text{off}}$ , and  $s(t+1) = s(t)$ ;
9   end
10 end
```

Suppose that M is the effective Markov matrix for the system when the ON/OFF policy in Algorithm 1 is active. The next theorem shows that M is a linear function of the design matrix K . This linearity property will be used for a convex synthesis of the algorithm design matrix K so that the matrix M satisfies some favorable properties as convergence and safety as we will show later on in the paper.

Theorem 1. Consider a system of single or multiple ON/OFF agents moving in a stochastic environment defined by a finite number of states \mathcal{S} with the transition probabilities given by G as in (3) for the ON mode and with no transitions for the OFF mode. Suppose that each agent executes the ON/OFF decision-making Algorithm 1 with the matrix K defined in (5). Then the density distribution $x(t)$ defined in (1) evolves based on a Markov chain as in (2), where the Markov matrix $M \in \mathbb{P}^{n \times n}$ is given as follows: For $i, j \in \mathbb{N}_n^+$,

$$M[i, j] = \begin{cases} G[i, j] K[i, j] & \text{if } i \neq j, \\ 1 - \sum_{\substack{k=1 \\ k \neq j}}^n G[k, j] K[k, j] & \text{if } i = j. \end{cases} \quad (6)$$

In matrix form, the above relationship is equivalent to

$$M = G \odot K + \text{diag}(\mathbf{1}^T - \mathbf{1}^T(G \odot K)). \quad (7)$$

Proof. Transition from the j^{th} state to the i^{th} state ($i \neq j$) can only take place when that transition is observed (line 4 in Algorithm 1) and accepted (line 6 in Algorithm 1):

$$\begin{aligned} M[i, j] &= \mathcal{P}\{s(t+1) = s_i | s(t) = s_j\} \\ &= \mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, y(t+1) = s_i | s(t) = s_j\} \end{aligned} \quad (8)$$

Using Bayes' Rule, we have:

$$\begin{aligned} M[i, j] &= \mathcal{P}\{y(t+1) = s_i | s(t) = s_j\} \times \\ &\quad \mathcal{P}\{\sigma(t) = \sigma_{\text{on}} | s(t) = s_j, y(t+1) = s_i\} \\ &= G[i, j] K[i, j] \end{aligned} \quad (9)$$

The system stays at the j^{th} state if either (a) the observed state at line 4 in Algorithm 1 is rejected (line 8 in Algorithm 1) or (b) the j^{th} state is observed at line 4 in Algorithm 1 and is accepted (line 6 in Algorithm 1), then

$$\begin{aligned} M[j, j] &= \mathcal{P}\{s(t+1) = s_j | s(t) = s_j\} \\ &= \mathcal{P}\{\sigma(t) = \sigma_{\text{off}} | s(t) = s_j\} \\ &\quad + \mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, y(t+1) = s_j | s(t) = s_j\} \\ &= 1 - \sum_{k=1}^n \mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, y(t+1) = s_k | s(t) = s_j\} \\ &\quad + \mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, y(t+1) = s_j | s(t) = s_j\} \\ &= 1 - \sum_{\substack{k=1 \\ k \neq j}}^n G[k, j] K[k, j] \end{aligned}$$

Note that we can verify that $M \in \mathbb{P}^{n \times n}$ by showing that $M \geq 0$ and $\mathbf{1}^T M = \mathbf{1}^T$.

$M \geq 0$: Letting $S := G \odot K$, since $0 \leq G$, $K \leq \mathbf{1}\mathbf{1}^T$, we have $\mathbf{1}^T S = \mathbf{1}^T (G \odot K) \leq \mathbf{1}^T G = \mathbf{1}^T$. Hence $\mathbf{1}^T - \mathbf{1}^T S \geq 0$, which implies that $M = S + \text{diag}(\mathbf{1}^T - \mathbf{1}^T S) \geq 0$.

$\mathbf{1}^T M = \mathbf{1}^T$: Let $\xi^T := \mathbf{1}^T G \odot K$. Then $\mathbf{1}^T M = \xi^T + \mathbf{1}^T \text{diag}(\mathbf{1}^T - \xi^T) = \xi^T + \mathbf{1}^T (I - \text{diag}(\xi)) = \xi^T + \mathbf{1}^T - \mathbf{1}^T \text{diag}(\xi) = \xi^T + \mathbf{1}^T - \xi^T = \mathbf{1}^T$. \square

Remark: The formulation for Markov matrix M given in (6) is quite intuitive. The probability of making transition $i \rightarrow j$ is simply the environment induced probability of this transition times the probability of accepting this transition. The diagonal entries of M are set so that the resulting Markov matrix satisfies the column stochasticity property. \blacksquare

IV. GENERALIZATION OF DENSITY CONTROL PROBLEM FOR MULTIPLE ON ACTIONS

In this section, we present the main result of this paper which is the generalization of the ON/OFF decision control policy problem for the case where the ON mode encapsulates multiple actions and the OFF mode features a single action that does not necessarily correspond to “no motion”. As explained earlier: in the ON mode, the “next step” outcomes of the actions are observable, while a Markov chain, G_{off} , is propagated when OFF mode is chosen. Now, we can have multiple actions in the ON mode whose transitions can be observed, i.e.,

$$\sigma(t) = \sigma_{on} \implies a(t) \in \mathcal{A}_{on} = \{a_1, \dots, a_m\}, \quad (10)$$

where $a(t)$ is the action taken in the ON mode. In the simple ON/OFF case of the previous section, we had $\mathcal{A}_{on} = \{a_1\}$, hence we did not need to explicitly define actions. Since we have multiple actions in the ON mode now, we have to explicitly identify them.

Using the same definitions for $x(t)$ and M given by (1) and (2), we will formulate the Markov matrix M as a function of the stochastic environment and the actions determined by a predetermined ON/OFF policy. For this general case, we expand the definitions of the probabilistic events for $t \in \mathbb{N}$:

$$\begin{aligned} y(t+1) = s_i &: \text{Observing a transition to state } s_i, \\ v(t) = a_k &: \text{Observing the outcome of taking action } \\ &\quad a_k \text{ from } \mathcal{A}_{on}, \\ a(t) = a_k &: \text{Accepting to execute action } a_k. \end{aligned}$$

Comparing to the previous model with single action, the addition here is the event “ $v(t) = a_k$ ”, which is used to define the probability of choosing an action whose outcome will be observed. The stochastic environment is defined with the transition matrices $G_k, G_{off} \in \mathbb{P}^{n \times n}$, $k \in \mathbb{N}_m^+$ where $G_k[i, j]$ gives the probability of observing a transition from j^{th} state to i^{th} state, given that the k^{th} action is selected to be observed, i.e.,

$$G_k[i, j] = \mathcal{P}\{y(t+1) = s_i | s(t) = s_j, v(t) = a_k\}, \quad (11)$$

where $i, j \in \mathbb{N}_n^+$, $k \in \mathbb{N}_m^+$ and similarly $G_{off}[i, j]$ defines the corresponding transition probabilities for the action in the OFF mode (e.g., $G_{off} = I$ when being OFF means no motion as in the previous section). The model for ON/OFF decision-making has the following assumptions (see Fig. 2):

- Each agent measures its own state at time instant t .
- Agent chooses a single action for the ON mode, say a_k , whose outcome will be observed, i.e., $v(t) = a_k$.
- Agent accepts or rejects to take the observed action.
- If action is accepted then it is taken, i.e., $a(t) = a_k$, and transitions occurs according to G_k .
- If action is rejected, the agent chooses the OFF mode, $\sigma(t) = \sigma_{off}$ and the transition occurs according to G_{off} .

Remark: In the above setting, G_{off} can have two interpretations: (i) the environment transition matrix when there is only one action for which the outcome is not observable; (ii) the effective transition matrix when there are multiple actions

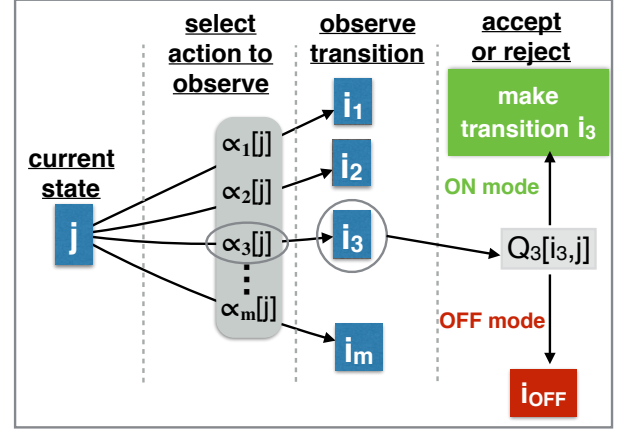


Fig. 2. Implementation of the decision policy.

with a prescribed decision policy. For the latter case, G_{off} is a Markov matrix defining the underlying Markov chain of an MDP with an existing policy running over the actions, which is discussed in more detail in the next subsection. ■

In the generalized model, we consider two sets of decision variables (to be designed offline):

$$\alpha_k[j] = \mathcal{P}\{v(t) = a_k | s(t) = s_j\} \quad (12)$$

$$Q_k[i, j] = \mathcal{P}\{a(t) = a_k | s(t) = s_j, v(t) = a_k, y(t+1) = s_i\} \quad (13)$$

with $k \in \mathbb{N}_m^+$, $i, j \in \mathbb{N}_n^+$. Namely, $\alpha_k[j]$ is the probability of choosing action $a_k \in \mathcal{A}_{on}$ at state s_j and $Q_k[i, j]$ is the probability of accepting an achievable transition $j \rightarrow i$ observed as an outcome of taking an action a_k .

Clearly Q_k matrices must be non-negative, $Q_k \in [0, 1]^{n \times n}$, $k \in \mathbb{N}_m^+$. Also, non-negative action variables α_k should satisfy the inequality $\sum_{k=1}^m \alpha_k[j] \leq 1$, $j \in \mathbb{N}_n^+$. It turns out in our model (which will become more clear later) that we can combine these two variables by considering the change of variables

$$P_k := Q_k \text{diag}(\alpha_k) = Q_k \odot (\mathbf{1} \alpha_k^T), \quad k \in \mathbb{N}_m^+. \quad (14)$$

The following property holds for P_k matrices,

$$P_k \leq \mathbf{1} \alpha_k^T, \quad k \in \mathbb{N}_m^+. \quad (15)$$

This inequality can simply be proven by contradiction. Suppose there exist i and j such that $P_k[i, j] > \alpha_k[j]$, then $Q_k[i, j] = P_k[i, j] / \alpha_k[j] > 1$ which is a contradiction because $Q_k[i, j] \in [0, 1]$.

We can now give by Algorithm 2 the ON/OFF decision-making policy for the general case. For this algorithm, we define variable $\phi \in \mathbb{R}^{m+2}$ where

$$\phi[1] = 0, \quad \phi[r] = \sum_{l=1}^r \alpha_l, \quad \phi[m+2] = 1$$

where $r = 2, \dots, m+1$.

The following theorem presents the key result in converting the ON/OFF decision policy design problem into a Markov chain synthesis problem.

Algorithm 2: ON/OFF Decision-Making Policy – General case

Inputs: $\{\alpha_k, Q_k : k \in \mathbb{N}_m^+\}$ (designed offline), \mathcal{S}, t_{\max}

```

1 for  $t \leftarrow 1$  to  $t_{\max}$  do
2   Determine current state  $s(t) \in \mathcal{S}$  (assume  $s(t) = s_j$ );
3   Generate a random numbers  $\mu(t) \sim U(0, 1)$  and  $\eta(t) \sim U(0, 1)$ ;
4   if  $\mu(t) \in [\phi[k], \phi[k+1]]$  then
5      $v(t) = a_k$ ;
6     Observe the next achievable transition for  $a_k$ :
7      $y(t+1)$  (suppose  $y(t+1) = s_i$ );
8     if  $\eta(t) \in [0, Q_k[i, j]]$  then
9       The agent switches to the mode ON,
10       $a(t) = a_k$  and  $s(t+1) = s_i$ ;
11    end
12  else
13    The agent switches to the mode OFF
14     $a(t) = a_{\text{off}}$ , and  $s(t+1)$  transitions according to  $G_{\text{off}}$ ;
15  end
16 end

```

Theorem 2. Consider a system of single or multiple mode switching ON/OFF agents moving in a stochastic environment defined by a finite number of states \mathcal{S} with the transition probabilities given by G_k as in (11) for the k^{th} action $a_k \in \mathcal{A}_{\text{on}}$ in the ON mode and G_{off} for the OFF mode. Suppose that each agent executes the ON/OFF decision-making Algorithm 2 with the matrices Q_k as in (13) and the vectors α_k as in (12). Then the state p.d.f. $x(t)$, defined in (1), evolves based on the Markov chain (2) with the Markov matrix $M \in \mathbb{P}^{n \times n}$ given by

$$M = \sum_{k=1}^m G_k \odot P_k + G_{\text{off}} \odot \left(\mathbf{1} \left(\mathbf{1}^T - \mathbf{1}^T \sum_{k=1}^m G_k \odot P_k \right) \right), \quad (16)$$

where $P_k, k \in \mathbb{N}_m^+$ are given by (14) and they satisfy

$$\sum_{k=1}^m \max_{i \in \mathbb{N}_n^+} P_k[i, j] \leq 1, \quad j \in \mathbb{N}_n^+. \quad (17)$$

Proof. The probability of making transition from j^{th} state to i^{th} state can be written as the following sum, since being ON and being OFF are mutually exclusive events:

$$\begin{aligned}
M[i, j] &= \mathcal{P}\{s(t+1) = s_i | s(t) = s_j\} \\
&= \underbrace{\mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, s(t+1) = s_i | s(t) = s_j\}}_{:=T_1} \\
&\quad + \underbrace{\mathcal{P}\{\sigma(t) = \sigma_{\text{off}}, s(t+1) = s_i | s(t) = s_j\}}_{:=T_2}.
\end{aligned} \quad (18)$$

In Algorithm 2, since executing the ON mode implies that an action from \mathcal{A}_{on} must be taken, the first term can be written

as the summation over all actions for the ON mode:

$$\begin{aligned}
T_1 &= \sum_{k=1}^m \mathcal{P}\{s(t+1) = s_i, a(t) = a_k | s(t) = s_j\} \\
&= \sum_{k=1}^m \mathcal{P}\{s(t+1) = s_i, a(t) = a_k, \\
&\quad v(t) = a_k | s(t) = s_j\}
\end{aligned}$$

The second equation above follows from the fact that $v(t)$ should always precedes $a(t)$, i.e., $\mathcal{P}(v(t) = a_k | a(t) = a_k) = 1$. By applying Bayes' rule [38] to the term inside the sum, we obtain:

$$\begin{aligned}
T_1 &= \sum_{k=1}^m \mathcal{P}\{s(t+1) = s_i, a(t) = a_k | v(t) = a_k, \\
&\quad s(t) = s_j\} \underbrace{\mathcal{P}\{v(t) = a_k | s(t) = s_j\}}_{\alpha_k[j]}.
\end{aligned}$$

Since observing transitions to distinct states are mutually exclusive events,

$$\begin{aligned}
T_1 &= \sum_{k=1}^m \left[\sum_{l=1}^n \mathcal{P}\{s(t+1) = s_i, a(t) = a_k, \right. \\
&\quad \left. y(t+1) = s_l | v(t) = a_k, s(t) = s_j\} \right] \alpha_k[j]
\end{aligned}$$

Applying Bayes' rule, we obtain:

$$\begin{aligned}
&= \sum_{k=1}^m \left[\sum_{l=1}^n \mathcal{P}\{s(t+1) = s_i, a(t) = a_k | \right. \\
&\quad \left. y(t+1) = s_l, v(t) = a_k, s(t) = s_j\} \right. \\
&\quad \times \underbrace{\mathcal{P}\{y(t+1) = s_l | v(t) = a_k, s(t) = s_j\}}_{G_k[l, j]} \left. \right] \alpha_k[j] \\
&= \sum_{k=1}^m \left[\sum_{l=1}^n \mathcal{P}\{s(t+1) = s_i | y(t+1) = s_l, \right. \\
&\quad \left. a(t) = a_k, v(t) = a_k, s(t) = s_j\} \right. \\
&\quad \times \underbrace{\mathcal{P}\{a(t) = a_k | y(t+1) = s_l, v(t) = a_k, s(t) = s_j\}}_{Q_k[l, j]} \\
&\quad \times G_k[l, j] \left. \right] \alpha_k[j]. \quad (19)
\end{aligned}$$

Note that, if the outcome of a_k is observed ($v(t) = a_k, y(t+1) = s_l$) and this action is accepted ($a(t) = a_k$), then the observed transition takes place with probability 1. In this case, a transition to any other state can take place only if the OFF mode is selected, i.e., $\sigma(t) = \sigma_{\text{off}}$. Hence, we have

$$\begin{aligned}
&\mathcal{P}\{s(t+1) = s_i | y(t+1) = s_l, a(t) = a_k, \\
&\quad v(t) = a_k, s(t) = s_j\} = \delta_{il},
\end{aligned} \quad (20)$$

where δ_{il} is the Kronecker delta operator, i.e., $\delta_{il} = 0$, if $i \neq l$ and $\delta_{il} = 1$, if $i = l$. We obtain the following equation by combining (19) and (20),

$$\begin{aligned}
T_1 &= \sum_{k=1}^m \left[\sum_{l=1}^n \delta_{il} Q_k[l, j] G[l, j] \right] \alpha_k[j] \\
&= \sum_{k=1}^m Q_k[i, j] G_k[i, j] \alpha_k[j].
\end{aligned} \quad (21)$$

Now, consider the second term in (18):

$$\begin{aligned} T_2 &= \mathcal{P}\{\sigma(t) = \sigma_{\text{off}}, s(t+1) = s_i | s(t) = s_j\} \\ &= \underbrace{\mathcal{P}\{s(t+1) = i | s(t) = s_j, \sigma(t) = \sigma_{\text{off}}\}}_{G_{\text{off}}[i,j]} \\ &\quad \times \underbrace{\mathcal{P}\{\sigma(t) = \sigma_{\text{off}} | s(t) = s_j\}}_{1 - \mathcal{P}\{\sigma(t) = \sigma_{\text{on}} | s(t) = s_j\}}. \end{aligned}$$

Here, given the current state, the probability of being ON is sum of the probabilities of taking each action for the ON mode, i.e.,

$$\begin{aligned} \mathcal{P}\{\sigma(t) = \sigma_{\text{on}} | s(t) = s_j\} &= \sum_{l=1}^n \mathcal{P}\{\sigma(t) = \sigma_{\text{on}}, s(t+1) = s_l | s(t) = s_j\} \\ &= \sum_{k=1}^m \sum_{l=1}^n Q_k[l, j] G_k[l, j] \alpha_k[j], \end{aligned}$$

which follows from (21). Now, combining the expressions we get for T_1 and T_2 , the equation (18) becomes:

$$\begin{aligned} M[i, j] &= \sum_{k=1}^m Q_k[i, j] G_k[i, j] \alpha_k[j] \\ &\quad + G_{\text{off}}[i, j] \left(1 - \sum_{k=1}^m \sum_{l=1}^n Q_k[l, j] G_k[l, j] \alpha_k[j] \right) \end{aligned}$$

which can be written in compact matrix notation as follows

$$\begin{aligned} M &= \sum_{k=1}^m G_k \odot Q_k \odot (\mathbf{1} \alpha_k^T) \\ &\quad + G_{\text{off}} \odot \left(\mathbf{1} \left(\mathbf{1}^T - \mathbf{1}^T \sum_{k=1}^m G_k \odot Q_k \odot (\mathbf{1} \alpha_k^T) \right) \right). \end{aligned} \quad (22)$$

With the change of variables given in (14), the equation given above is equivalent to (16).

Finally, we will show that $M \in \mathbb{P}^{n \times n}$. For nonnegativity, we will consider M in two terms. As both terms corresponds to probabilistic quantities, they both must be nonnegative. The nonnegativity of the first term, $\sum_{k=1}^m G_k \odot P_k$ is clear since $G_k \geq 0$ and $P_k \geq 0$, $k \in \mathbb{N}_m^+$. For the nonnegativity of the second term, i.e.,

$$G_{\text{off}} \odot \left(\mathbf{1} \left(\mathbf{1}^T - \mathbf{1}^T \sum_{k=1}^m G_k \odot P_k \right) \right)$$

we need to show

$$\mathbf{1}^T \sum_{k=1}^m G_k \odot P_k \leq \mathbf{1}^T. \quad (23)$$

Note that $\mathbf{1}^T \sum_{k=1}^m G_k \odot P_k$ can be written in index notation as follows: For $j \in \mathbb{N}_m^+$,

$$\mathbf{1}^T \left(\sum_{k=1}^m G_k \odot P_k \right) e_j = \sum_{i=1}^n \sum_{k=1}^m G_k[i, j] P_k[i, j]. \quad (24)$$

Here, $\sum_{k=1}^m P_k[i, j] \leq 1$ for any i and j since

$$\sum_{k=1}^m P_k \leq \sum_{k=1}^m \mathbf{1} \alpha_k^T \leq \mathbf{1} \mathbf{1}^T,$$

and $\sum_{i=1}^n G_k[i, j] = 1$ for any j and k since G_k is column stochastic. Hence,

$$\begin{aligned} \sum_{i=1}^n \sum_{k=1}^m G_k[i, j] P_k[i, j] &= \sum_{k=1}^m \text{conv} P_k[i, j] \\ &\leq \sum_{k=1}^m \max_{i \in \mathbb{N}_n^+} P_k[i, j] \\ &\leq 1. \end{aligned}$$

The last inequality follows from (17), which then implies that

$$\mathbf{1}^T \left(\sum_{k=1}^m G_k \odot P_k \right) e_j \leq 1, \quad j \in \mathbb{N}_n^+,$$

and hence (23) holds. This concludes that $M \geq 0$.

With $G_{\text{off}} \in \mathbb{P}^{n \times n}$, the above inequality implies that

$$G_{\text{off}} \odot \left(\mathbf{1} \left(\mathbf{1}^T - \mathbf{1}^T \sum_{k=1}^m G_k \odot P_k \right) \right) \geq 0,$$

and hence $M \geq 0$. Next let $H := \sum_{k=1}^m G_k \odot P_k$, then

$$\begin{aligned} \mathbf{1}^T M &= \mathbf{1}^T H + \mathbf{1}^T G_{\text{off}} \odot (\mathbf{1} (\mathbf{1}^T - \mathbf{1}^T H)) \\ &= \mathbf{1}^T H + \mathbf{1}^T (G_{\text{off}} - G_{\text{off}} \odot (\mathbf{1} \mathbf{1}^T H)) \\ &= \mathbf{1}^T H + \mathbf{1}^T - \mathbf{1}^T (G_{\text{off}} \odot (\mathbf{1} \mathbf{1}^T H)). \end{aligned}$$

Let $\phi := H^T \mathbf{1}$, then

$$\begin{aligned} \mathbf{1}^T M &= \phi^T + \mathbf{1}^T - \mathbf{1}^T (G_{\text{off}} \odot (\mathbf{1} \phi^T)) \\ &= \phi^T + \mathbf{1}^T - (\mathbf{1}^T G_{\text{off}}) \odot \phi^T \\ &= \phi^T + \mathbf{1}^T - \mathbf{1}^T \odot \phi^T = \phi^T + \mathbf{1}^T - \phi^T = \mathbf{1}^T, \end{aligned}$$

hence $M \in \mathbb{P}^{n \times n}$. \square

Theorem 2 shows that M is a linear function of $\{P_k : k = 1, \dots, m\}$. This linearity property will be used for a convex synthesis of P_k so that the matrix M satisfies some favorable properties as convergence and safety as we will show later on in the paper. The algorithm design parameters α_k and Q_k can then be extracted from P_k as we will show next. Thus the design of $\{P_k : k = 1, \dots, m\}$ is an intermediary step to set the parameters of Algorithm 2.

A. Extraction of α_k and Q_k from P_k

Once P_k matrices are computed (which will be explained in later sections), the selection of α_k and Q_k can be done in multiple ways, that is, the same P_k matrices can be parameterized in multiple ways via α_k and Q_k . The choice must preserve the following conditions on α_k and Q_k (since they contain probabilities of events as entries)

$$0 \leq Q_k \leq \mathbf{1} \mathbf{1}^T, \quad 0 \leq \alpha_k \leq \mathbf{1}, \quad k \in \mathbb{N}_m^+, \quad \sum_{k=1}^m \alpha_k \leq \mathbf{1}. \quad (25)$$

Our default parameterization is:

$$\alpha_k[j] = \max_{i \in \mathbb{N}_n^+} P_k[i, j], \quad k \in \mathbb{N}_m^+, \quad j \in \mathbb{N}_n^+. \quad (26)$$

Hence the last inequality in (25) is satisfied (due to (17)), and we can choose Q_k as

$$Q_k = P_k \text{diag}(\alpha_k)^{-1}, \quad k \in \mathbb{N}_m^+. \quad (27)$$

constraints for $t \in \mathbb{N}$:

Transition :

$$\mathcal{P}\{s(t+1)=s_i|s(t)=s_i\}=0 \text{ if } A_a[j,i]=0, \quad (29)$$

Safety :

$$Lx(t) \leq p, \quad \forall x(t) \leq q, \quad (30)$$

Convergence :

$$\lim_{t \rightarrow \infty} x(t) = v, \quad \forall x(0) \in \mathbb{P}^n, \quad (31)$$

where v is a desired, discrete, probability distribution, and A_a is the adjacency matrix defining the allowable transitions between states between two consecutive time steps, L , q , and p are given matrices and arrays specifying safety constraints. In this section, we provide brief discussions on these constraints and express them as equivalent convex constraints on the Markov matrix M . Since the matrix M depends linearly on the ON/OFF decision policy matrices P_k as in (16), this will imply that the constraints will be convex constraints on P_k 's, which are our design variables. Once P_k 's are computed, we can choose α_k 's and Q_k 's by using (26) or (28) and (27).

Transition Constraints:

It is useful to impose constraints on the physically realizable state transitions by specifying some entries of the matrix M as zeros. In the case of ON/OFF agents, these constraints are automatically ensured by stochastic transition matrices: If a state transition is simply not possible naturally, then the corresponding entry in the matrix G_k , $\forall k$ is zero, which implies that it is also zero in the matrix M , i.e., $M[i,j] = 0$ if $G_k[i,j] = 0, \forall k$. If additional transitions are also needed to be eliminated due to mission specific reasons, we can impose additional constraints beyond the ones that are naturally imposed by matrix G with the following constraint:

$$(\mathbf{1}\mathbf{1}^T - A_a^T) \odot M = \mathbf{0} \quad (32)$$

where A_a is the adjacency matrix, i.e., $A_a[i,j] = 1$, if transition $i \rightarrow j$ is allowable and $A_a[i,j] = 0$ otherwise. Note that, this constraint is already linear, hence convex in M .

Safety Constraints:

We consider a general form of linear density safety constraints:

$$Lx(t) \leq p, \quad \forall x(t) \leq q \quad (33)$$

which covers several types of safety constraints with different selections of L , p and q . Two examples of these constraints are: (i) Density upper bound constraints; (ii) Density rate constraints. Density upper bound constraint ensures that the density of each state stays below a prescribed value, that is,

$$x(t) \leq d, \quad t = 1, 2, \dots \quad (34)$$

where $\mathbf{0} \leq d \leq \mathbf{1}$ defines the density upper bounds in each state and it is assumed that $x(0) \leq d$. Note that, this form can be obtained by letting $L = M$, $q = d$ and $p = d$ in the general form given in (33). The density rate constraint is used to limit the rate of change of density for each state of Markov chain.

$$-f \leq x(t+1) - x(t) \leq f, \quad t = 0, 1, \dots, \quad (35)$$

where $f \geq \mathbf{0}$ bounds the flow rate. Note that this is equivalent to case where

$$L = \begin{bmatrix} M - I \\ I - M \end{bmatrix}, \quad q = \begin{bmatrix} f \\ f \end{bmatrix},$$

and $p = d$ in (33). For more examples on the linear safety constraints captured by the general form and for the proof of Lemma 1, see [9].

The safety constraints given in (33) are ensured by the following lemma, which gives necessary and sufficient conditions for safety as linear inequalities on M .

Lemma 1. [39], [9] Consider the Markov chain given by (2). Then,

$$Lx(t) \leq q, \quad \forall x(t) \leq p, \quad (36)$$

if and only if there exist $S \in \mathbb{R}^{n \times n}$ and $y \in \mathbb{R}^n$ such that

$$\begin{aligned} S &\geq \mathbf{0}, \quad [L + S + y\mathbf{1}^T] \geq \mathbf{0}, \\ y + q &\geq [L + S + y\mathbf{1}^T]p. \end{aligned} \quad (37)$$

So far, we have obtained the linear equivalent conditions on Markov matrix for the transition, safety and ON/OFF constraints. Hence, for the convex optimization problem, we define a set for feasible Markov matrices which satisfy safety, transition and ON/OFF constraints:

$$\mathcal{M}_F = \{M \in \mathbb{P}^{n \times n} : M \text{ satisfies (16), (17), (32), (37)}\}.$$

$M \in \mathcal{M}_F$ if and only if:

$$M \geq \mathbf{0}, S \geq \mathbf{0}, P_k \geq \mathbf{0} \text{ for } k = 1 \dots m,$$

$$\mathbf{1}^T M = \mathbf{1}^T,$$

$$(\mathbf{1}\mathbf{1}^T - A_a^T) \odot M = \mathbf{0},$$

$$[L + S + y\mathbf{1}^T] \geq \mathbf{0},$$

$$y + q \geq [L + S + y\mathbf{1}^T]p,$$

$$M = \sum_{k=1}^m G_k \odot P_k$$

$$+ G_{\text{off}} \odot \left(\mathbf{1} \left(\mathbf{1}^T - \mathbf{1}^T \sum_{k=1}^m G_k \odot P_k \right) \right),$$

$$\sum_{k=1}^m \max_{i \in \mathbb{N}_n^+} P_k[i, j] \leq 1, \quad j \in \mathbb{N}_n^+$$

(38)

Note that the last inequality above in (38) can be written by using linear inequalities, hence it is a convex constraint. To see that define

$$Z_j := [P_1(:, j) \ P_2(:, j) \ \dots \ P_m(:, j)], \quad j = 1, \dots, n,$$

where $P_k(:, j)$ is the j 'th column of P_k . Then we can replace the last inequality by the following inequalities

$$Z_j \leq \mathbf{1}\beta_j^T, \quad \mathbf{1}^T \beta_j \leq 1, \quad j = 1, \dots, n,$$

where β_j 's are $m \times 1$ slack variables.

Formulation of Convergence/Coverage Constraints:

As a part of the mission objectives for multi-agent systems, the agent distribution $x(t)$ is required to converge to a desired distribution $v \in \mathbb{P}^n$, i.e.,

$$\lim_{t \rightarrow \infty} x(t) = v \quad \forall x(0) \in \mathbb{P}^n. \quad (39)$$

Since M is a Markov matrix, hence column stochastic, a necessary condition for the desired convergence is that the desired distribution $v \in \mathbb{P}^n$ is an eigenvector of M :

$$Mv = v. \quad (40)$$

Notice that $Mv = v$ are simple linear equalities, and hence they are convex constraints but are not sufficient for convergence. In order to obtain the convex equivalent of the ergodicity constraint ($\lim_{t \rightarrow \infty} x(t) = v \quad \forall x(0) \in \mathbb{P}^n$), there are three possible approaches. The first approach is to use the results in [3] which developed a sufficient condition for ergodicity based on the LMI characterizations of the stability of discrete-time systems (by leveraging results from [40]). The sufficient condition is : There exist $P = P^T \succ \mathbf{0}$, $\lambda \in [0, 1)$, and F such that

$$\begin{bmatrix} \lambda^2 P & (M - v\mathbf{1}^T)^T F^T \\ F(M - v\mathbf{1}^T) & F + F^T - P \end{bmatrix} \succeq 0. \quad (41)$$

The above condition is an LMI for prescribed values of λ and F which we typically choose to be $F = \text{diag}(v)^{-1}$. A more detailed discussion on the selection of F and λ can be found in [39]. Here, λ denotes the exponential decay rate of the error, $e(t) = x(t) - v$ and can be computed via a line search (by noting that the above inequalities are LMIs for a given λ and $F = \text{diag}(v)^{-1}$) to achieve the largest feasible convergence rate.

As a second approach, we can impose *reversibility* on the Markov matrix. Then, if $v > \mathbf{0}$, we can state a necessary and sufficient condition [41] for ergodicity constraint as follows:

$$-\lambda I \preceq H^{-1}MH - hh^T \preceq \lambda I, \quad (42)$$

where $h = (v_1^{1/2}, \dots, v_m^{1/2})$ and $H = \text{diag}(h)$. Note that λ , the convergence rate, can be minimized for fastest convergence within a convex problem formulation.

Finally, the third approach is to impose strong connectivity on the corresponding adjacency matrix of the Markov matrix, i.e. $\mathbf{i}(M)$ is strongly connected. This can be obtained using the following linear equations:

$$M[i, j] \geq \epsilon \text{ if } A_a[i, j] = 1, \quad (43)$$

where $\epsilon > 0$ is a small positive scalar (e.g., it can be set as the machine precision). This provides a sufficient condition for ergodicity of the Markov chain. Thus if $Mv = v$ and the adjacency matrix A_a corresponds to a connected graph, then v is unique stationary distribution and $\lim_{t \rightarrow \infty} x(t) = v \quad \forall x(0) \in \mathbb{P}^n$. Note that, though we can ensure convergence with simpler inequalities, we cannot impose or minimize convergence rate directly in this approach.

Formulation of synthesis as an optimization problem:

We can formulate the Markov matrix synthesis as a minimization problem with the desired constraints. One example cost function is $\mathbf{1}^T(\mathbf{1} - \text{diag}(M))$ which aims to minimize overall action, i.e., $M \simeq I$:

$$\begin{aligned} \min_M \quad & \mathbf{1}^T(\mathbf{1} - \text{diag}(M)) \quad \text{such that} \\ & M \in \mathcal{M}_F \\ & (41) \text{ or } (42) \text{ or } (43). \end{aligned} \quad (44)$$

VI. NUMERICAL EXAMPLE

This section presents an illustrative numerical example for the density control problem for autonomous agents with ON/OFF control modes. We consider a swarm of mobile agents that are distributed over the configuration space (see Figure 4) that is partitioned to 8 subregions, which are referred to as *bins*. In this configuration, probabilistic density distribution is given as $x[i](t) := \mathcal{P}\{r(t) \in R_i\}$ where $r(t)$ is the position vector of an agent at time step t . For this setting, safety upper bound constraint is used to limit the expected number of agents in each bin. Two sets of simulations are performed by using the ON/OFF policy synthesized by solving the LMI problem in (44) with (41), both with the density upper bound constraints and $m = 5$ actions for the ON mode: (i) Total $N_{s1} = 3000$ simulations with same safe initial condition, i.e., same $x(0)$ with different realizations; (ii) Total $N_{s2} = 3000$ simulations with randomly generated 3000 safe initial conditions. For all cases, OFF case corresponds to “no action”, i.e., $G_{\text{off}} = I$. For the actions in the ON mode, column stochastic G_k matrices are selected such that they have different steady-state final distributions and do not satisfy safety and transition constraints.

Other parameters for the simulations are set as follows:

$$A_a = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$N_a = 3000$$

$$x(0) = [0.5 \ 0 \ 0.5 \ 0 \ 0 \ 0 \ 0 \ 0]^T$$

$$v = [0.005 \ 0.02 \ 0.005 \ 0.04 \ 0.05 \ 0.34 \ 0.2 \ 0.34]^T$$

$$d = [1 \ 0.15 \ 1 \ 0.12 \ 0.12 \ 1 \ 0.4 \ 1]^T$$

$$\lambda = 0.975$$

where A_a is the adjacency matrix of the bin connections, N_a is the total number of agents, $x(0)$ is the initial distribution of agents, v is the desired final distribution as in equation (31), d is a safety upper bound constraint (as in equation (33) with $L = I$ and $p = q = d$), and λ is the convergence rate of the system.

Since the bins at the corners behave like accumulation points in the initial and final distributions, the safety upper bound constraints are not imposed for these bins, i.e., the

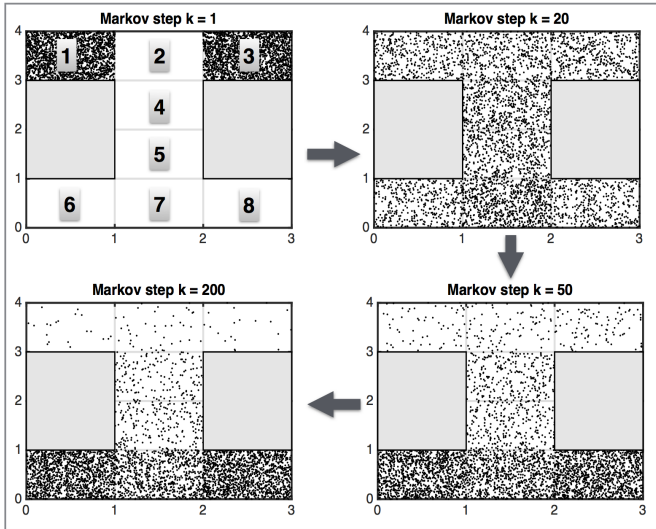


Fig. 4. Snapshots of simulation: configuration space with bin numbers

corresponding entries of the d vector are set as 1. With the given parameters, the optimization problem given in (38) with (41) is solved using YALMIP and SDPT3 [42], [43]. G_k matrices used in the simulations and the final distributions, v_k , corresponding to each G_k are given in the appendix along with the resulting solution variables Q_k and α_k . In many applications, environmental transition matrices, G_k 's, may satisfy transition constraints in most examples, i.e., $G_k[i, j] = 0$ when $A_a[j, i] = 0$. However this example considers some G_k matrices that do not have this property, that is, we do not allow some motions even when they can be induced by the environment. For example, as given in the appendix, $G_1[4, 1] > 0$ even though $A_a[1, 4] = 0$ and hence $Q_1[4, 1] = 0$. This makes the problem more challenging since motion constraints are not automatically satisfied by the environment and they must be ensured by the policy. Such scenarios can arise in the balloon motion control example given in the introduction [13], [12], [5] where environmental transition matrices may not satisfy the desired constraints. Though some altitudes may induce high velocities, we may not choose to ride with such fast winds, for example, not to damage the structural integrity of the balloon (there may be a maximum speed limit for structural safety purposes).

Simulation results are presented in Figure 5. The mean density \bar{x} and 3σ confidence bounds are shown for the case with density upper bound d . The average density for the case without constraints is obtained by evolving the density according to equation (2). Density goes above the desired upper bound for the bins 2, 4 and 5 when the constraint is not imposed. By using ON/OFF control policy, we are able to ensure that the density does not go beyond the prescribed upper limit at a reasonable cost of reduced convergence rate.

For the results of the second set of simulations, point-wise maximum values of the density at each time step over all 3000 simulations are plotted. This is a good demonstration of our claim in Lemma 1: For all safe initial conditions, i.e., $x(0) \leq d$, the density is guaranteed to satisfy safety constraints for all time, i.e., $x(t) \leq d$, $t \in \mathbb{N}^+$.

The snapshots of the overall distribution taken at the beginning and at the end of a simulation from the first set are shown in Figure 4 which also shows the bin numbers.

VII. CONCLUSION

In this paper, we develop a probabilistic density control policy for autonomous mobile agents with two modes: ON and OFF. When the agent is in the ON mode, it can observe the one-step outcome of a single action chosen from actions for the ON mode and can decide whether to take this action or not. If it does not take the action, it switches to the OFF mode. The density distribution of agents in the system evolves according to a Markov chain that, as shown in this paper, is a linear function of the stochastic environment and the decision policy. We formulate a convex optimization problem, that can be solved reliably via interior-point methods, to synthesize the decision policy which ensures desired safety, transition and convergence properties for the underlying Markov chain. The given constraints on the density are equivalently expressed as constraints on the Markov chain. The resulting density control model is illustrated with a numerical example on autonomous mobile agents.

REFERENCES

- [1] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, "System architecture directions for networked sensors," *ACM SIGOPS operating systems review*, vol. 34, no. 5, pp. 93–104, 2000.
- [2] C. M. Saaj, V. Lappas, and V. Gazi, "Spacecraft swarm navigation and control using artificial potential field and sliding mode control," in *Industrial Technology, 2006. ICIT 2006. IEEE International Conference on*. IEEE, 2006, pp. 2646–2651.
- [3] B. Açıkmese and D. S. Bayard, "Markov chain approach to probabilistic guidance for swarms of autonomous agents," *Asian Journal of Control*, vol. 17, no. 4, pp. 1105–1124, 2015.
- [4] P. Corke, S. Hrabar, R. Peterson, D. Rus, S. Saripalli, and G. Sukhatme, "Autonomous deployment and repair of a sensor network using an unmanned aerial vehicle," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 4. IEEE, 2004, pp. 3602–3608.
- [5] M. T. Wolf, L. Blackmore, Y. Kuwata, N. Fathpour, A. Elfes, and C. Newman, "Probabilistic motion planning of balloons in strong, uncertain wind fields," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1123–1129.
- [6] J. Heidemann, W. Ye, J. Wills, A. Syed, and Y. Li, "Research challenges and applications for underwater sensor networking," in *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006*, vol. 1. IEEE, 2006, pp. 228–235.
- [7] K. Avrachenkov, M. E. Chamie, and G. Neglia, "A local average consensus algorithm for wireless sensor networks," in *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, June 2011, pp. 1–6.
- [8] B. Açıkmese and D. S. Bayard, "A markov chain approach to probabilistic swarm guidance," *American Control Conference, Montreal, Canada*, pp. 6300–6307, 2012.
- [9] N. Demir, U. Eren, and B. Açıkmese, "Decentralized probabilistic density control of autonomous swarms with safety constraints," *Autonomous Robots*, vol. 39, no. 4, pp. 537–554, 2015.
- [10] N. Demir and B. Açıkmese, "Probabilistic density control for swarm of decentralized on-off agents with safety constraints," *American Control Conference (ACC)*, 2015, pp. 5238–5244, 2015.
- [11] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [12] Y. Kuwata, L. Blackmore, M. Wolf, N. Fathpour, C. Newman, and A. Elfes, "Decomposition algorithm for global reachability analysis on a time-varying graph with an application to planetary exploration," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3955–3960.

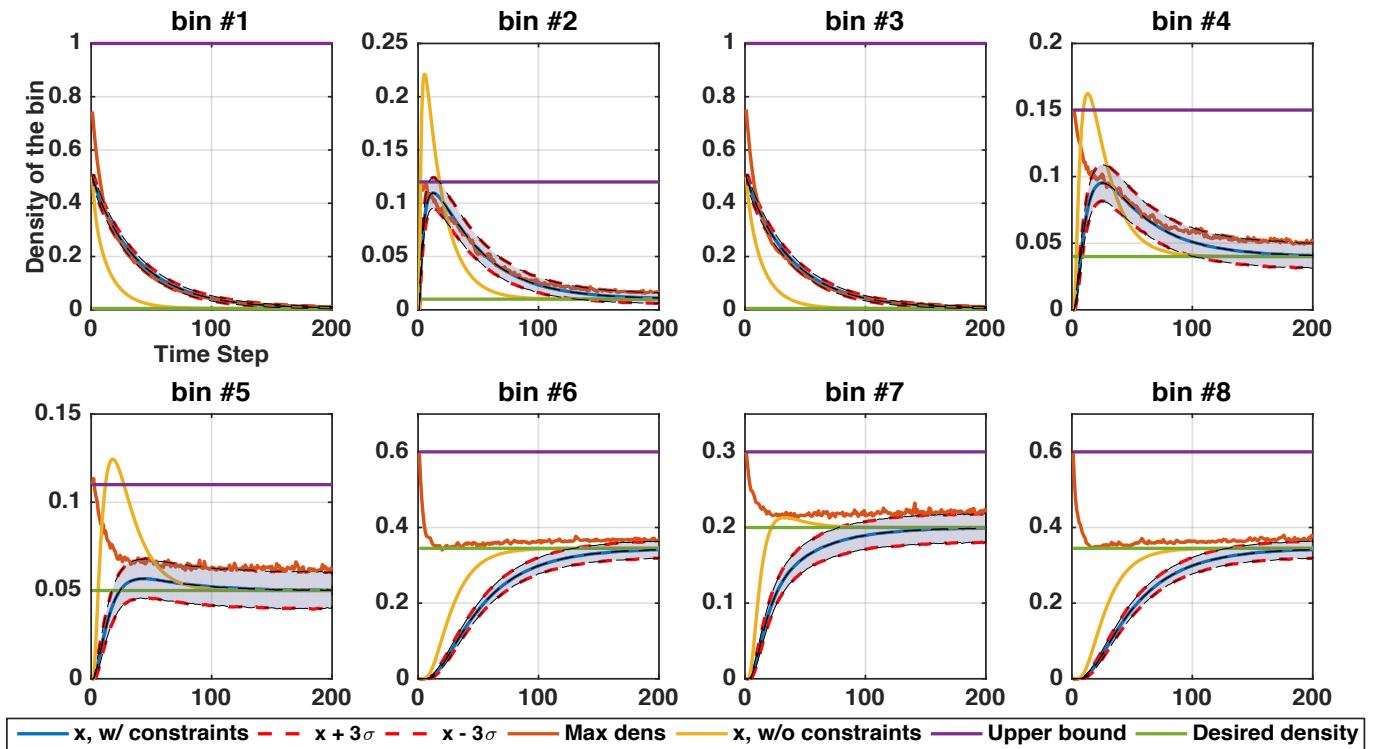


Fig. 5. Time history of the density of each bin. With the density upper bound constraints the density is guaranteed to stay in the tube between red dashed lines with 99.7% confidence.

- [13] A. Elfes, K. Reh, P. Beauchamp, N. Fathpour, L. Blackmore, C. Newman, Y. Kuwata, M. Wolf, and C. Assad, "Implications of wind-assisted aerial navigation for titan mission planning and science exploration," in *Aerospace Conference, 2010 IEEE*. IEEE, 2010, pp. 1–7.
- [14] O. Sigaud and O. Buffet, *Markov Decision Processes in Artificial Intelligence*. Wiley-IEEE Press, 2010.
- [15] R. Zhang, Y. Yu, M. El Chamie, B. Açıkmeşe, and D. Ballard, "Decision-making policies for heterogeneous autonomous multi-agent systems with safety constraints," in *25th International Joint Conference on Artificial Intelligence (IJCAI-16)*, July 2016.
- [16] M. El Chamie, Y. Yu, and B. Açıkmeşe, "Convex synthesis of randomized policies for controlled Markov processes with hard safety constraints," in *2016 American Control Conference (ACC)*, July 2016.
- [17] M. El Chamie and B. Açıkmeşe, "Convex synthesis of optimal policies for markov decision processes with sequentially-observed transitions," in *2016 American Control Conference (ACC)*, July 2016.
- [18] C. Amato, G. Chowdhary, A. Geramifard, N. K. Üre, and M. J. Kochenderfer, "Decentralized control of partially observable markov decision processes," in *52nd IEEE Conference on Decision and Control*. IEEE, 2013, pp. 2398–2405.
- [19] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," *Mathematics of Operations Research*, vol. 27, no. 4, pp. 819–840, 2002. [Online]. Available: <http://rbr.cs.umass.edu/shlomo/papers/BGIZmor02.html>
- [20] S. Berman, A. Halasz, M. A. Hsieh, and V. Kumar, "Optimized stochastic policies for task allocation in swarms of robots," *IEEE Trans. on Robotics*, vol. 25, no. 4, pp. 927–937, 2009.
- [21] I. Chattopadhyay and A. Ray, "Supervised self-organization of homogeneous swarms using ergodic projections of markov chains," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 39, no. 6, pp. 1505–1515, 2009.
- [22] M. Lawford and W. Wonham, "Supervisory control of probabilistic discrete event systems," *Proc. 36th Midwest Symp. Circuits Syst.*, pp. 327–331, 1993.
- [23] A. R. Mesquita, J. P. Hespanha, and K. Astrom, "Optimotaxis: A stochastic multi-agent on site optimization procedure," *Hybrid Systems: Computation and Control, Lecture Notes in Computational Science*, no. 4981, pp. 358–371, 2008.
- [24] A. Jadbabaie, G. J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [25] D. P. Bertsekas and J. N. Tsitsiklis, "Comments on on coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. on Automatic Control*, vol. 52, no. 5, pp. 968–969, 2007.
- [26] S. Martínez, F. Bullo, J. Cortés, and E. Frazzoli, "On synchronous robotic networks-part ii: Time complexity of rendezvous and deployment algorithms," *IEEE Trans. on Automatic Control*, vol. 52, no. 12, pp. 2214–2226, 2007.
- [27] S. Roy, "Scaled consensus," *Automatica*, vol. 51, pp. 259–262, 2015.
- [28] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt, "Controllability of multi-agent systems from a graph-theoretic perspective," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 162–186, 2009.
- [29] J. Cortés, S. Martínez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.
- [30] M. Pavone, E. Frazzoli, and F. Bullo, "Adaptive and distributed algorithms for vehicle routing in a stochastic and dynamic environment," *IEEE Trans. on Automatic Control*, vol. 56, no. 6, pp. 1259–1274, 2011.
- [31] H. Bai, M. Arcak, and J. T. Wen, "Adaptive design for reference velocity recovery in motion coordination," *Systems & control letters*, vol. 57, no. 8, pp. 602–610, 2008.
- [32] M. A. Hsieh, V. Kumar, and L. Chaimowicz, "Decentralized controllers for shape generation with robotic swarms," *Robotica*, vol. 26, no. 5, pp. 691–701, 2008.
- [33] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: A game-theoretical formulation," *ASME Jnl. of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 584–197, 2007.
- [34] V. J. Lumelsky and K. R. Harinarayan, "Decentralized motion planning for multiple mobile robots: The cocktail party model," *Autonomous Robots*, vol. 4, no. 1, pp. 121–135, 1997.
- [35] M. Pavone and E. Frazzoli, "Decentralized policies for geometric pattern formation and path coverage," *ASME Jnl. of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 633–643, 2007.
- [36] M. A. Hsieh, A. Halasz, S. Berman, and V. Kumar, "Biologically inspired redistribution of a swarm of robots among multiple sites," *Swarm Intelligence*, vol. 2, no. 2-4, pp. 121–141, 2008.
- [37] S. Martínez, J. Cortés, and F. Bullo, "Motion coordination with distributed information," *IEEE Control System Magazine*, pp. 75–88, 2007.

- [38] K. L. Chung, *A Course in Probability Theory*. Academic Press, 2001.
- [39] B. Açıkmese, N. Demir, B. Açıkmese, and M. Harris, "Convex necessary and sufficient conditions for density safety constraints in markov chain synthesis," *IEEE Trans. on Automatic Control*, vol. 60, no. 10, pp. 2813–2818, 2015.
- [40] M. de Oliveira, J. Bernussou, and J. Geromel, "Anew discrete-time robust stability condition," *Systems and Control Letters*, pp. 261–265, 1999.
- [41] S. Boyd, P. Diaconis, P. Parillo, and L. Xiao, "Fastest mixing Markov Chain on graphs with symmetries," *SIAM Jnl. of Optimization*, vol. 20, no. 2, pp. 792–819, 2009.
- [42] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: <http://control.ee.ethz.ch/~jloef/yalmip.php>
- [43] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using sdpt3," *Mathematical Programming*, vol. 95, no. 2, pp. 189–217, 2003.

APPENDIX

The problem parameters and the resulting solution variables for the numerical example are:

$$G_1 = 10^{-4} \begin{bmatrix} 6735 & 384 & 284 & 146 & 179 & 27 & 34 & 22 \\ 773 & 5252 & 290 & 339 & 337 & 24 & 42 & 22 \\ 680 & 1080 & 7611 & 245 & 141 & 116 & 135 & 83 \\ 505 & 1196 & 226 & 6428 & 278 & 95 & 436 & 122 \\ 202 & 891 & 389 & 613 & 6425 & 79 & 421 & 88 \\ 391 & 308 & 585 & 450 & 546 & 8012 & 506 & 1264 \\ 349 & 540 & 157 & 1339 & 1696 & 666 & 7437 & 379 \\ 365 & 349 & 458 & 440 & 398 & 981 & 989 & 8020 \end{bmatrix}$$

$$G_2 = 10^{-4} \begin{bmatrix} 1871 & 2359 & 1563 & 1590 & 1349 & 2650 & 2471 & 235 \\ 1616 & 82 & 1651 & 2019 & 1008 & 56 & 1159 & 1820 \\ 1269 & 194 & 283 & 1387 & 1117 & 924 & 2903 & 1992 \\ 1314 & 2350 & 2205 & 149 & 1586 & 522 & 402 & 1211 \\ 1209 & 695 & 1657 & 1766 & 1792 & 1928 & 101 & 1641 \\ 824 & 614 & 1911 & 1827 & 511 & 2070 & 1478 & 1032 \\ 1463 & 1806 & 155 & 778 & 1038 & 111 & 992 & 453 \\ 434 & 1900 & 575 & 484 & 1599 & 1739 & 494 & 1616 \end{bmatrix}$$

$$G_3 = 10^{-4} \begin{bmatrix} 4579 & 389 & 324 & 291 & 360 & 53 & 69 & 46 \\ 975 & 3954 & 359 & 293 & 320 & 47 & 84 & 43 \\ 822 & 1304 & 5688 & 490 & 282 & 233 & 270 & 167 \\ 1009 & 1171 & 451 & 5333 & 269 & 190 & 336 & 244 \\ 404 & 787 & 777 & 750 & 5460 & 158 & 301 & 175 \\ 783 & 617 & 1171 & 901 & 1092 & 7854 & 1012 & 697 \\ 699 & 1081 & 314 & 1062 & 1422 & 878 & 7083 & 432 \\ 729 & 697 & 916 & 880 & 795 & 587 & 845 & 8196 \end{bmatrix}$$

$$G_4 = 10^{-4} \begin{bmatrix} 5864 & 878 & 526 & 686 & 679 & 764 & 70 & 534 \\ 780 & 4938 & 783 & 671 & 516 & 768 & 799 & 744 \\ 571 & 906 & 4983 & 689 & 638 & 1027 & 761 & 424 \\ 795 & 1231 & 693 & 5033 & 461 & 124 & 849 & 814 \\ 687 & 771 & 593 & 925 & 4788 & 751 & 873 & 612 \\ 744 & 560 & 836 & 665 & 598 & 4581 & 951 & 1066 \\ 388 & 276 & 572 & 804 & 1530 & 807 & 4885 & 738 \\ 171 & 440 & 1014 & 527 & 790 & 1178 & 812 & 5068 \end{bmatrix}$$

$$G_5 = 10^{-4} \begin{bmatrix} 8890 & 377 & 244 & 0 & 0 & 0 & 0 & 0 \\ 572 & 6550 & 221 & 385 & 354 & 0 & 0 & 0 \\ 538 & 856 & 9535 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1221 & 0 & 7522 & 287 & 0 & 535 & 0 \\ 0 & 996 & 0 & 477 & 7390 & 0 & 540 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8170 & 0 & 1830 \\ 0 & 0 & 0 & 1616 & 1969 & 455 & 7792 & 326 \\ 0 & 0 & 0 & 0 & 0 & 1375 & 1133 & 7844 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 0.0200 \\ 0.0200 \\ 0.0600 \\ 0.0600 \\ 0.0600 \\ 0.2900 \\ 0.2000 \\ 0.2900 \end{bmatrix}, v_2 = \begin{bmatrix} 0.1756 \\ 0.1191 \\ 0.1197 \\ 0.1246 \\ 0.1391 \\ 0.1246 \\ 0.0883 \\ 0.1090 \end{bmatrix}, v_3 = \begin{bmatrix} 0.0200 \\ 0.0200 \\ 0.0600 \\ 0.0600 \\ 0.0600 \\ 0.2900 \\ 0.2000 \\ 0.2900 \end{bmatrix},$$

$$v_4 = \begin{bmatrix} 0.1250 \\ 0.1250 \\ 0.1250 \\ 0.1250 \\ 0.1250 \\ 0.1250 \\ 0.1250 \\ 0.1250 \end{bmatrix}, v_5 = \begin{bmatrix} 0.0200 \\ 0.0200 \\ 0.0600 \\ 0.0600 \\ 0.0600 \\ 0.2900 \\ 0.2000 \\ 0.2900 \end{bmatrix}.$$

$$\alpha_1 = \begin{bmatrix} 0.0940 \\ 0.1139 \\ 0.0568 \\ 0.2006 \\ 0.1975 \\ 0.1656 \\ 0.1573 \\ 0.1477 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 0.5594 \\ 0.5165 \\ 0.6630 \\ 0.1332 \\ 0.1125 \\ 0.1229 \\ 0.2167 \\ 0.1553 \end{bmatrix}, \alpha_3 = \begin{bmatrix} 0.1201 \\ 0.1036 \\ 0.0616 \\ 0.1457 \\ 0.1492 \\ 0.1846 \\ 0.2036 \\ 0.1488 \end{bmatrix},$$

$$\alpha_4 = \begin{bmatrix} 0.0941 \\ 0.0987 \\ 0.0969 \\ 0.1135 \\ 0.1676 \\ 0.1845 \\ 0.2150 \\ 0.1955 \end{bmatrix}, \alpha_5 = \begin{bmatrix} 0.0976 \\ 0.1464 \\ 0.0679 \\ 0.3711 \\ 0.3404 \\ 0.1896 \\ 0.1537 \\ 0.1903 \end{bmatrix}$$

$$Q_1 = 10^{-4} \begin{bmatrix} 6184 & 6205 & 7025 & 0 & 0 & 0 & 0 & 0 \\ 10000 & 5983 & 10000 & 4886 & 7151 & 0 & 0 & 0 \\ 7165 & 9178 & 7084 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 5888 & 4599 & 0 & 6142 & 0 \\ 0 & 9932 & 0 & 8553 & 5895 & 0 & 7780 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6645 & 9431 & 8970 \\ 0 & 0 & 0 & 10000 & 10000 & 10000 & 6141 & 10000 \\ 0 & 0 & 0 & 0 & 0 & 8186 & 10000 & 7044 \end{bmatrix}$$

$$Q_2 = 10^{-4} \begin{bmatrix} 5788 & 4625 & 5634 & 0 & 0 & 0 & 0 & 0 \\ 10000 & 5780 & 10000 & 2621 & 8247 & 0 & 0 & 0 \\ 9130 & 8424 & 5778 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 6133 & 2381 & 0 & 5928 & 0 \\ 0 & 9913 & 0 & 9595 & 6170 & 0 & 6557 & 0 \\ 0 & 0 & 0 & 0 & 0 & 7643 & 10000 & 8495 \\ 0 & 0 & 0 & 10000 & 10000 & 8729 & 5970 & 10000 \\ 0 & 0 & 0 & 0 & 0 & 10000 & 9441 & 6867 \end{bmatrix}$$

$$Q_3 = 10^{-4} \begin{bmatrix} 6037 & 6198 & 6769 & 0 & 0 & 0 & 0 & 0 \\ 10000 & 6009 & 10000 & 5376 & 7016 & 0 & 0 & 0 \\ 7375 & 9295 & 6838 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 6006 & 4978 & 0 & 6045 & 0 \\ 0 & 9875 & 0 & 8594 & 5989 & 0 & 7564 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6406 & 10000 & 8110 \\ 0 & 0 & 0 & 10000 & 10000 & 10000 & 6069 & 10000 \\ 0 & 0 & 0 & 0 & 0 & 7443 & 9937 & 6929 \end{bmatrix}$$

$$Q_4 = 10^{-4} \begin{bmatrix} 6181 & 6542 & 6026 & 0 & 0 & 0 & 0 & 0 \\ 10000 & 6010 & 10000 & 4975 & 7535 & 0 & 0 & 0 \\ 7022 & 8884 & 6154 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 6170 & 4073 & 0 & 5961 & 0 \\ 0 & 9840 & 0 & 8807 & 5947 & 0 & 8956 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6455 & 10000 & 8130 \\ 0 & 0 & 0 & 10000 & 10000 & 10000 & 6073 & 10000 \\ 0 & 0 & 0 & 0 & 0 & 8266 & 9948 & 6375 \end{bmatrix}$$

$$Q_5 = 10^{-4} \begin{bmatrix} 6326 & 6227 & 7107 & 0 & 0 & 0 & 0 & 0 \\ 10000 & 5957 & 10000 & 3764 & 7591 & 0 & 0 & 0 \\ 7150 & 9133 & 7195 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10000 & 0 & 5857 & 3727 & 0 & 6123 & 0 \\ 0 & 9974 & 0 & 8794 & 5868 & 0 & 8017 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6913 & 6109 & 9742 \\ 0 & 0 & 0 & 10000 & 10000 & 10000 & 6134 & 10000 \\ 0 & 0 & 0 & 0 & 0 & 9090 & 10000 & 7037 \end{bmatrix}$$



Nazlı Demirer is a Graduate Research Assistant with the Autonomous Control Laboratory in the William E. Boeing Department of Aeronautics and Astronautics at the University of Washington, USA. She received the B.S. degree in Aerospace Engineering from the Middle East Technical University, Ankara, Turkey in 2012 and M.S degree from The University of Texas at Austin, TX in 2014. She is currently pursuing Ph.D. degree in Aeronautics and Astronautics from the University of Washington. Her current research interests include autonomous multi-

agent systems and Markov chains.



Mahmoud El Chamie is a Research Associate with the Autonomous Control Laboratory in the William E. Boeing Department of Aeronautics and Astronautics at the University of Washington, USA. He received a M.S. (UBINET program) and a Ph.D. in computer science from the University of Nice Sophia Antipolis, France, in 2011 and 2014 respectively. His Ph.D. was held at Inria Sophia Antipolis with the Maestro team. His current research interests include optimization and control of autonomous multi-agent systems.



Behçet Açıkmese is an Associate Professor in the Department of Aeronautics and Astronautics at the University of Washington, Seattle. He received his Ph.D. from Purdue University and was a Visiting Assistant Professor at Purdue University before joining NASA Jet Propulsion Laboratory (JPL) in 2003. He was a senior technologist at JPL and a lecturer at Caltech, where he developed control algorithms for planetary landing, formation flying spacecraft, and asteroid and comet sample return missions. He is the developer of the flyaway control algorithms used in

Mars Science Laboratory, which landed on Mars in August, 2012. He has joined the faculty of University of Texas at Austin in 2012, where he stayed until 2015.